

引用格式:彭锦辉,张功凯,王彤,等.轨迹驱动的多层时空图神经网络交通路况短期预测[J].地球信息科学学报,2024,26(10):2300-2315.
[Peng J H, Zhang G K, Wang T, et al. A trajectory-driven multi-layer spatiotemporal graph neural network for predicting short-term urban traffic state[J]. Journal of Geo-information Science, 2024,26(10):2300-2315.] DOI:10.12082/dqxxkx.2024.240175
CSTR:32074.14.dqxxkx.2024.240175

轨迹驱动的多层时空图神经网络交通路况短期预测

彭锦辉¹, 张功凯¹, 王 彤¹, 王培晓², 张 彤^{1*}

1. 武汉大学 测绘遥感信息工程国家重点实验室,武汉 430070; 2. 中国科学院地理科学与资源研究所 资源与环境信息系统国家重点实验室,北京 100101

A Trajectory-Driven Multi-Layer Spatiotemporal Graph Neural Network for Predicting Short-Term Urban Traffic State

PENG Jinhui¹, ZHANG Gongkai¹, WANG Tong¹, WANG Peixiao², ZHANG Tong^{1*}

1. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430070, China; 2. State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research , Chinese Academy of Sciences, Beijing 100101, China

Abstract: The short-term prediction of urban traffic states serves as a fundamental application supporting traffic management and online navigation. Floating car trajectory data, characterized by low cost and high spatio-temporal coverage, has been widely employed for real-time traffic state extraction, thereby facilitating short-term traffic prediction. However, the spatio-temporal coverage of floating car trajectories is extremely uneven, resulting in a significant amount of missing or insufficient coverage of road networks over many time periods. It is difficult to accurately estimate the traffic states of all segments of the entire road network directly using the trajectory data of floating cars. The accuracy and reliability cannot meet the needs of real-time traffic state estimation and short-term prediction. This issue therefore poses a significant technical challenge for online short-term prediction of traffic states based on uneven trajectory data across the entire road network in large cities, hindering the refinement of traffic monitoring and management. To address the issue of uneven spatiotemporal distribution in trajectory data, this study proposes a dynamic layering method for the road network, dividing it into multiple layers based on the spatio-temporal distribution of trajectories, including a main road network with high-quality trajectories and other secondary road networks with sparse trajectory distributions. Building upon the layered road network, we propose a trajectory-driven multi-level spatio-temporal graph neural network method for short-term traffic state prediction. Leveraging the hierarchical structure of road networks, the proposed prediction method incorporates intra- and inter-layer message passing mechanisms that consider the spatio-temporal distribution of trajectories. We apply dilated causal convolution and graph attentional mechanisms to describe the complex spatio-temporal correlations of traffic states among road networks. Based on the correlation representation, we develop a unified graph neural network prediction model that integrates "representation" and "prediction" into an end-to-end learning scheme. The proposed method enables online prediction of speed and

收稿日期:2024-03-30;修回日期:2024-05-23.

基金项目:武大-华为空间技术创新实验室2023年开放基金项目。 [Foundation item: Open Funds of the Wuhan University - Huawei Geoinformatics Innovation Laboratory.]

作者简介:彭锦辉(2001—),男,江西吉安人,硕士生,主要从事交通预测研究。E-mail: jh_peng@qq.com

*通讯作者:张 彤(1979—),男,福建武夷山人,博士,教授,主要从事时空机器学习、高分遥感解译、交通地理信息系统研究。

E-mail: zhangt@whu.edu.cn

congestion state of all road segment in the road network simultaneously, significantly enhancing the predictive performance of traffic state for road segments with sparse trajectories. Through testing on real trajectory data from the large road network of Wuhan City, the proposed method showcases a notable improvement in prediction accuracy compared to recent popular baseline methods, particularly achieving satisfactory performance in segments with severe trajectory data deficiencies. The training efficiency is also significantly improved. Experimental results indicate that the proposed multi-level spatio-temporal graph neural network prediction method can effectively handle the prediction challenges caused by the uneven distribution of trajectories.

Key words: traffic state; short-term prediction; multi-layer spatiotemporal graph neural network; trajectory; road network partition; correlation representation of traffic state; message passing

*Corresponding author: ZHANG Tong, E-mail: zhangt@whu.edu.cn

摘要:城市交通路况的短期预测是支持交通管理、在线导航的基础应用。出租车轨迹作为低成本、高时空覆盖率的交通监测数据类型,已广泛用于提取实时路况,支持路况短期预测。然而出租车轨迹时空覆盖极不均衡,导致大量的路段和时段轨迹数据缺失或覆盖率不足,难以直接基于轨迹数据精确估计全路网所有路段全天候的交通状态,精度和可靠性都不能满足实时交通路况估计和短期预测的需要。因而基于不均衡轨迹数据的全路网交通路况在线短期预测成为大城市交通精细化监测和管理的一大技术难题。本文针对轨迹数据的时空分布不均衡问题,设计了路网的动态分层方法,将城市路网根据轨迹的时空分布划分为多层次路网,包括轨迹质量较好的主干路网以及轨迹分布较为稀疏的次级路网。在分层路网基础上,我们提出轨迹驱动的多层次时空图神经网络路况短期预测方法,依托不同路网层级建立多层次时空图神经网络,设计顾及轨迹时空分布的层内和层间消息传递机制,基于因果膨胀卷积和图注意力描述路网之间复杂的路况时空关联。在路况关联表征模型基础上,设计实现了表征-预测一体化的集成端到端图神经网络预测模型,可同时对全路网所有路段的速度和状态进行在线预测,有效提升轨迹分布稀疏路段的路况预测质量。通过武汉市大型路网的真实轨迹数据测试,本方法比基线方法在预测精度上有显著改善,特别是在轨迹数据缺失较严重的路段上取得了较好的预测性能,同时训练效率也有显著提升,表明所提出的多层次时空图神经网络预测方法能有效地应对轨迹分布不均衡导致的路况预测难题。

关键词:交通路况;短期预测;多层次时空图神经网络;轨迹;路网分层;路况关联表征;消息传递

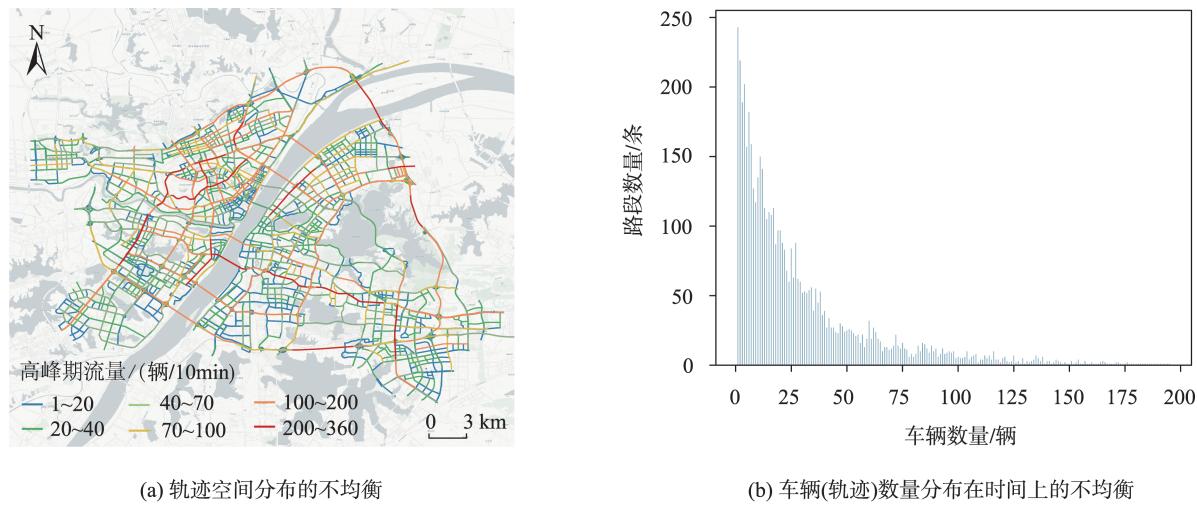
1 引言

城市交通路况(Traffic State)指的是给定时间和空间位置的流量、速度、占有率等场景交通参数,有时也涉及排队长度、行程时间等变量以及交通拥堵的分析。在实际交通管理和导航等应用中,路况信息主要指的是通信的速度和拥堵状态2个方面。城市交通路况的短期预测是支持交通管理、在线导航的基础应用,也已经在主流网络地图服务中得到广泛部署。为了支持路况短期预测,需要获取可靠的城市实时路况信息。当前业界主要依托大规模浮动车特别是出租车轨迹数据,融合线圈、视频等多源交通监测信息提取路况^[1-2]。浮动车轨迹作为低成本、高时空覆盖率的交通监测数据类型,已广泛用于提取实时路况^[3],支持路况短期预测。

然而出租车轨迹时空覆盖极不均衡,空间上主干道路车流量较大,但是许多次要道路或新建道路的浮动车覆盖率较低。从图1(a)的流量统计和图1(b)中的统计直方图观察而知,车辆数量在不同路段之间呈现较大的数量差异。时间上,高峰期

和平峰期的轨迹覆盖也有很大差异。轨迹数据的时空分布不均衡问题导致大量的路段和时段轨迹数据缺失或覆盖率不足,难以直接基于轨迹数据精确估计这些时段下全路网所有路段的交通状态,精度和可靠性都不能满足实时交通路况估计和预测的需要。在轨迹覆盖率较低的区域,交通状态参数可能被低估^[4]。较大城市路网环境复杂,轨迹数据覆盖情况的不均衡情况更加突出。顾及轨迹数据时空不均衡,实现全路网交通路况的在线短期预测因而成为大城市交通精细化监测和管理的一大技术难题。基于大规模出租车轨迹数据,针对轨迹数据时空分布不均衡问题,采用先进的时空机器学习方法,深入挖掘和表征路网中交通路况的时空关联和变化模式,对于提升城市交通路况的预测精度和效率,改进城市交通在线控制和组织都有着重要的应用意义。

认识到传统的统计参数模型在先验预设的严格限制以及在非线性非平稳方面描述能力的不足,智能交通系统已经广泛采纳了基于机器学习的路况短期预测方法,用于对未来半小时到一小时内的路网交通路况或参数进行在线预测^[5-7],从而更好地



注：以武汉市2023年某工作日早高峰8:00—8:10的流量统计为例，轨迹在空间上分布极不均衡(图(a))。路段在同一时段的车辆数量分布差异显著。

图1 轨迹分布的不均衡示意
Fig. 1 Uneven distribution of trajectories

支持交通组织和管理。由于交通短期预测问题涉及时空2个维度上关联的建模，常见的方法将所有路段所有时间内的路况数据看作多元时间序列，使用循环神经网络及其变体门控循环单元和长短期记忆网络提取时间维度特征，使用卷积神经网络或残差网路提取空间维度特征，然后将时空维度特征融合起来进行预测^[8-9]。近年来，由于其对网络表征的强大能力，图神经网络能够捕获交通领域中常见的非欧几里得结构数据中隐藏的空间信息，图神经网络开始大量应用于交通短期预测问题^[7,10]。研究人员对基础图神经网络进行定制和扩展，使之能提取路网中交通时序数据的时空关联特征，典型的工作包括时空图卷积网络(ST-GCN)^[11]、图因果膨胀卷积网络(Graph Wavenet)^[12]、时序图卷积网络(T-GCN)^[13]和残差图卷积长短期记忆网络(RGC-LSTM)^[14]等。常规图神经网络多数在层数较少的情况下才能取得较好的预测性能，过多的神经网络层会导致过平滑问题^[15]，影响了图神经网络的表征性能，因此在交通路网中难以从高级邻域中提取到远程的时空关联信息。部分研究尝试在建模中集成宏观和微观层次的交通信息，如通过图结构学习获取自适应邻接矩阵，更好地表征多尺度交通时空关联^[16]，或通过构建层级图卷积网络来融合多层次交通关联信息^[17]。最近研究进一步在各种先进图神经网络的基础上进行定制，关注对交通路况动态时空关联的精确描述，如动态图卷积循环网络(DGCRN)^[18]、动态Transformer模型(PDFomer)^[19]

和因果图时空预测网络(Causal-GCNM)^[20]等。以上这些研究多数直接基于已预先提取的路况信息进行建模和推理，未能考虑到轨迹数据在时空上分布的不均衡，因此所预测的路况许多轨迹分布稀疏路段可能精度和可靠性都较低。

交通短期预测建模依赖于各种监测设备采集的交通参数信息，这些信息往往带有缺失问题。在应对缺失数据方面，多数经典的深度学习预测方法没有显式地考虑数据缺失^[11-14]，只是简单地认为训练和测试数据在时空维度上都是完整。也有方法只进行简单的预处理补全缺失部分，甚至干脆忽视缺失部分。这些策略不能正确恢复交通模式，显然会对后续预测性能带来负面影响^[21]，因此良好的数据补全策略是提升交通短期预测的重要环节^[22]。本研究顾及了轨迹数据稀疏带来的路段交通信息缺失问题，通过合理的分层建模策略和缺失预补全方法，减少了数据缺失对预测性能的影响。

因此，近年来研究人员提出了一些考虑缺失的交通预测方法，在现有的深度学习网络中加入特定的模块或者利用矩阵或者张量分解方法，从而更好地应对数据缺失问题^[21,23-26]。特别是出现了显式顾及缺失模式图神经网络预测方法，能更合理地描述和恢复路网的动态时空关联，因此相对于传统不考虑缺失的预测方法在精度和稳定性方面都有一定的提高。如Wang等^[24]提出一种动态时间图神经网络，使用Traffic BERT学习动态邻接矩阵，再集成图神经网络和门控循环单元学习缺失场景下交通流

时空特征。Cui等^[26]提出基于谱马尔可夫网络的越策方法,将马尔可夫过程结合谱图卷积运算,拓展至图结构,逐步推断交通状态实现预测。但是这种策略如果应用到大型路网,由于需要对所有路段同时进行表征建模,计算复杂度较高,不能很好地满足在线预测任务以及模型的迭代更新需要。此外,顾及缺失的预测模型认为特定路段和时段的交通数据是全部缺失,也不适用于轨迹数据稀疏分布的路段,即稀疏数据的信息需要适当利用而不能当做全部缺失。本研究同样在图神经网络框架下建立显式考虑数据缺失的交通短期预测模型,与上述文献不同的是,通过对路网分层,将缺失较为严重的路段与其他主干路段分离开,通过层间消息传递方法将主干路网提取的路况时空信息传递至轨迹稀疏的次级路网,实现交通路况信息的在线补全和快速预测。实验证明,分层建模方法具备更快的训练效率,有利于模型的更新和大规模路网的建模。

目前大多数研究是关注高速无信控路网或小规模路网的交通预测问题,网络的复杂度较低,节点数量局限在百级,少有研究关注城市大型信控路网,并显性地表征其层次结构。面向大城市信控路网,基于大规模轨迹数据、顾及轨迹时空分布与质量的实时交通路况提取和预测的解决方案还比较缺乏。当前学术界和产业界所提出和采用的各种方法,包括传统分析建模、机器学习建模方法都没有很好地针对轨迹数据时空分布不均衡的问题提出高效、精确而可靠的交通路况预测方法。本文针对轨迹数据的时空分布不均衡问题,设计了路网的动态分层方法,将城市路网根据轨迹的时空分布划分为多级路网,包括轨迹质量较好的主干路网以及轨迹分布较为稀疏的次级或三级路网。主干路网,描述长期相对稳定的路况时空关联,而其他等级路网受到轨迹数据的限制,路况信息较不可靠,需要结合主干路网的路况来描述路况的时空关联。因此,在分层路网基础上,提出轨迹驱动的多层时空图神经网络路况预测方法(A Trajectory-driven Multi-level Spatio-Temporal Graph Neural Network for Traffic State Prediction, T-MGNN),依托不同路网的层级建立多层时空图神经网络,设计顾及轨迹时空分布的层内和层间消息传递机制,较好地提升了城市大型路网的交通路况预测性能。通过武汉大型路网的真实轨迹数据测试,本方法比近年来流行的基线方法在预测精度上有显著改善,特别是在轨迹数据缺失较严重的路段上取得了较好的预测性能。

2 研究方法

2.1 问题定义

城市路网可以抽象为图结构 $G = \langle V, E, A \rangle$, 其中 $V = \{v_i\}_{i=1}^n$ 表示 n 条路段的集合, E 为图的边集合, 表示路段之间的拓扑连接。 $A \in \mathbb{R}^{n \times n}$ 是描述路段连接的邻接矩阵, 如果两条路段相连, 邻接矩阵对应位置设置为 1, 反之为 0。路网所有路段在一定时间范围 T 的路况观测值为 $X \in \mathbb{R}^{n \times T}$, $x^t = \{x_i^t\}_{i=1}^n \in \mathbb{R}^{n \times 1}$ 表示所有路段在第 t 个时刻记录的路况数据, 其中 x_i^t 表示路段 v_i 在第 t 个时刻记录的路况数据(速度、拥堵状况等)。由于本研究主要的观测数据来源是车辆轨迹, 因此路况观测数据是基于轨迹数据经过地图匹配、速度提取等步骤得到。

为了进行交通短期预测, 需要建立一个可学习参数 θ 的函数模型 F , 使用 p 步历史观测值 $X_{t-p}^t = \{x^{\tau}\}_{\tau=t-p}^t \in \mathbb{R}^{n \times p}$ 预测未来 q 步的路况数据 $\hat{X}_{t+1}^{t+q} \in \mathbb{R}^{n \times q}$, 公式如下:

$$\hat{X}_{t+1}^{t+q} = \{\hat{x}^{t+1}, \hat{x}^{t+2}, \dots, \hat{x}^{t+q}\} = F_{\theta}\{X_{t-p}^t; G\} \quad (1)$$

式中: F_{θ} 即为所提出的多层时空图神经网络路况预测模型。需要注意的是, 本研究预测的路况包括未来路网所有路段的平均速度以及相应的交通拥堵状态。

2.2 轨迹驱动的多层图时空神经网络预测模型总体框架

本研究提出一种轨迹驱动的多层时空图神经网络路况预测方法, 针对轨迹时空分布的不均衡问题, 建立了路网的分层方法以及相对应的多层时空图神经网络的交通路况表征框架, 有效地提升轨迹分布稀疏路段的路况补全和预测质量, 设计了面向主次路网的消息传递机制和自适应关联建模方法, 支持全面可靠地描述路段之间的复杂交通时空关联。图 2 展示了方法的整体技术框架和流程, 主要包括 4 个主要模块, 即数据处理、路况表征建模、模型训练以及模型测试或应用。数据处理模块主要进行数据预处理、路网分层以及训练/测试数据集的构建, 面向特定城市, 采集整理路网和一定时间范围的出租车轨迹数据, 通过地图匹配等预处理将轨迹数据关联到路网。根据轨迹数据的时空分布模式将路网分成 2—3 级, 并构造各种矩阵作为分层路网信息。面向每个路段, 利用轨迹点包含的速度信

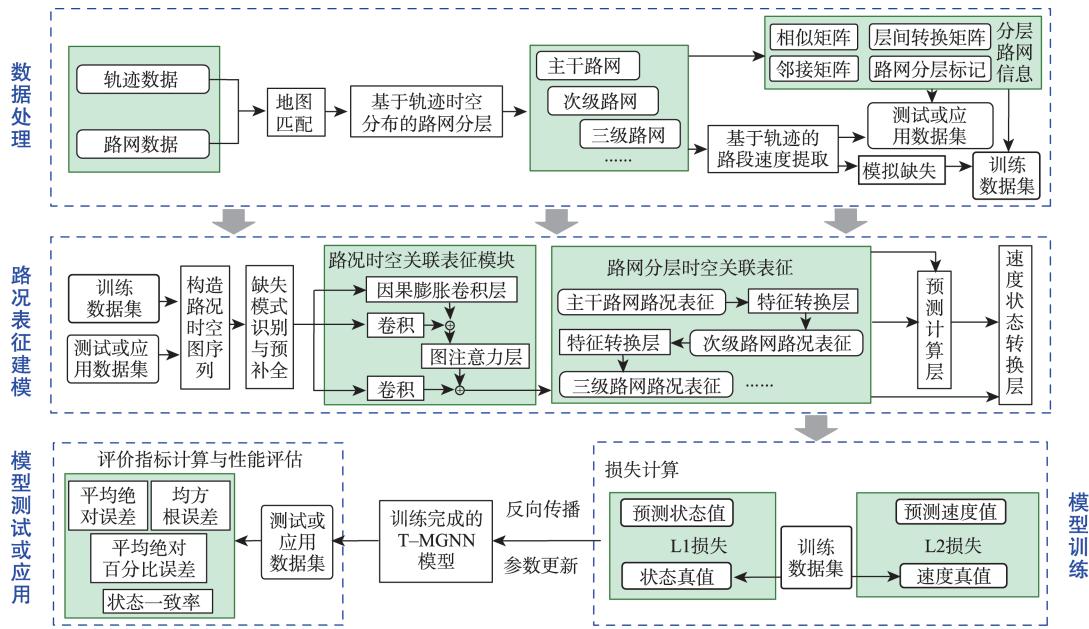


图2 轨迹驱动的多层时空图神经网络交通路况短期预测技术框架和流程

Fig. 2 Technical framework and process of trajectory-driven multi-layer spatiotemporal graph neural network for short-term traffic condition prediction technology

息,去除异常值后,设计实现速度提取算法计算每条路段每个时间窗口(如2~5 min)的平均速度。对于无轨迹点或者轨迹点过于稀疏的路段,无法准确提取平均速度信息,则进行标记。将速度信息处理后转化为路段路况,完成测试和训练数据集构造。

数据表征模块则是依托训练或测试数据集,设计基于多层图神经网络的分层路况时空关联表征模块描述路段之间复杂的路况关联和依赖关系。模型训练模块则主要面向速度和拥堵状态构造损失函数,利用训练数据集对模型进行训练。训练完成后,可基于测试或应用数据集对模型进行测试和应用。

本方法涵盖了从原始轨迹数据到全路网交通路况的预测完整实用化技术流程,建立了表征-预测一体化的集成多层图神经网络模型,实现了直接基于原始轨迹数据进行大型路网的路况在线预测。

2.3 面向轨迹时空分布的路网分层

根据出租车轨迹的实际时空分布对路网进行动态分层可以将区分轨迹质量差异较大的路段,通过路网分层可以基于不同层的路网直接构建多层图神经网络预测模型。首先我们采用一种结合隐马尔可夫模型和预计算的快速地图匹配算法^[23]将原始轨迹数据匹配到路网。轨迹数据时空分布不平衡导致路网中不同路段的轨迹分布存在明显的差异,城市的主要道路由于车流量较大,轨迹数据

分布密集,提取的速度数据质量较高。而次要和分支路段轨迹分布稀疏,提取的速度数据质量不佳。根据轨迹数据的时空分布模式将路网所有的路段分成若干层(一般为2—3层)(图3)。其中,主干路网承担大部分的交通流量,是整个城市交通的基本骨架,绝大部分时段都能从轨迹中提取较为完整可靠的路况信息。此外主干路网应保持整体上的连通和完整。其他层级路网则可能存在大量轨迹数据稀疏的路段,车辆密度随着时段波动较大,这类路段上监测到的速度可靠性相对较低。由于轨迹数据的限制,划分三层路网会造成第三层路网轨迹数量过于稀疏,因此本文主要将研究区域的路网分为两层进行预测建模,即主干路网和次级路网。后续实验也按照两层路网进行测试。

本文采用基于流量统计的方法进行分层,取特定时段的轨迹数据(如工作日早高峰),统计每个路段预设时间粒度(如5 min)的平均流量,并绘制流量的概率分布函数,设置预设阈值,通过流量数据分布进行路网分层。一般可取流量在前20%的路段作为主干路网。设路段流量范围是 $(0, q_{\max})$, $P(q)$ 代表路段流量的概率密度函数: $P(q) = \int_0^{q_{\max}} p(q) dq = 1$, 有 $P(0 \leq q \leq q_{\max}) = \int_0^{q_{\max}} p(q) dq = 1$, 其中 q_{\max} 为路段最大平均流量。若选择前20%的路段作为主干路网,则应求得使 $P(q_{0.8})=0.8$ 时 $q_{0.8}$ 的值,

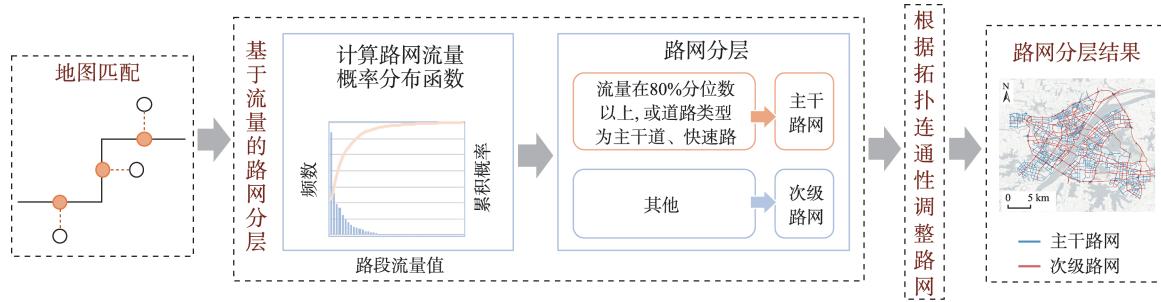


图3 路网分层方法

Fig. 3 Road network layering method

平均流量大于等于 $q_{0.8}$ 的路段则为主干路网。其他分层由此类推,按照一定的百分位确定次级以及更低等级的路网。再根据拓扑连通性调整路网,保整主干路网整体连通和完整性。分层后,主干路网多数路段在大部分时段都能直接从轨迹中提取较完整可靠的路况信息,次等级路网存在大量轨迹数据稀疏的路段。如果轨迹数据分布极不均衡,大量路段存在轨迹缺失,则可考虑基于缺失率进行分层。基于一定时期(如一周或者一个月)的历史轨迹数据,可结合轨迹数据实际的时空分布特点,针对工作日、周末、高峰和非高峰等不同时段提取不同的分层路网。

在分层完成后,每个路段记录其所属层的标记,同时需要构建多个矩阵描述分层路网信息,为后续表征和预测建模提供基础数据:

(1)建立路段之间的相似性矩阵。把各路段的速度值时序数据 $x_i = \{x_i^t\}_{t=1}^T \in \mathbb{R}^{1 \times T}$ 作为向量,计算路段向量之间的欧式距离 $d_{ij} = \sqrt{\sum_{t=1}^T (x_i^t - x_j^t)^2}$,作为各路段之间的相似性,按相似度递减排序的前 z 条路段作为邻域(缺省设置为15条),建立相似性矩阵 A ,距离越小说明相似度越大;

(2)建立不同层路网之间的转换矩阵。层间转换矩阵 S 依据主干路网和次级路网的拓扑关系,大小为 $c_1 \times c_2$, c_1 为主干路网数量, c_2 为次级路段数量。转换矩阵各行 s_i 代表各主干路段,各列 s_j 代表各次级路段,如果不同层路网的路段相连,相应矩阵位置 s_{ij} 赋值为1,否则为0。

2.4 基于分层路网的路况关联表征与预测

在完成路网分层后,可以在不同路网上进行路况的关联表征建模,以路段为基本单元,基于历史路况数据(包括平均速度和拥堵状态),相应地设计

多层次时空图神经网络描述路段之间路况的复杂时空关联。通过分层,可以先对较高路况质量的主干路网进行表征建模,然后通过层间消息传递机制在次级路网上有效集成主干网络的表征信息,从而有效提升轨迹分布稀疏路段的路况表征和预测质量。通过面向主次路网的消息传递机制和自适应关联建模方法,支持全面可靠地描述路段之间的复杂交通时空关联。

图4以两层路网为例描述了分层路网关联表征的流程。对于不同层的路网,根据基准数据集(训练或测试/应用数据集)、路网邻接矩阵、相似性矩阵和分层路网转换矩阵,构造路况时序图。每个时段,每个层级路网构造一个路况时空图,路况数据记录为时空状态矩阵 $X \in R^{n \times d}$,标记缺失路况信息的路段和时段。不同时段的路网图以及相应的时空状态矩阵可构造为路况时空图序列,为建立路况表征模型提供基础数据结构。然后如果路段路况数据质量较低,存在大量缺失情况,需要对时序数据进行预补全,提高模型的路况表征能力。这里采用的方法是一种感知缺失模式的在线时空线性插值方法,算法细节可参见^[20]。预补全后,分别应用多个路况时空关联表征模块(Spatio-Temporal CorrelationExtractor, STCE),该模块集成因果膨胀卷积层和图注意力层提取路况时序关联特征和空间关联特征,即层内路况的关联表征建模。路况关联表征模型同时学习速度和状态特征。在完成各层层内的路况关联表征后,由于次级路网的表征是基于低质量轨迹数据,因此需要集成主干路网的表征。本文定义了一种层间消息传递机制,主干路网层的信息经过特征转换层单向流动至次级路网层,从而将主干路网提取的路况信息传递至轨迹稀疏的次级路网。

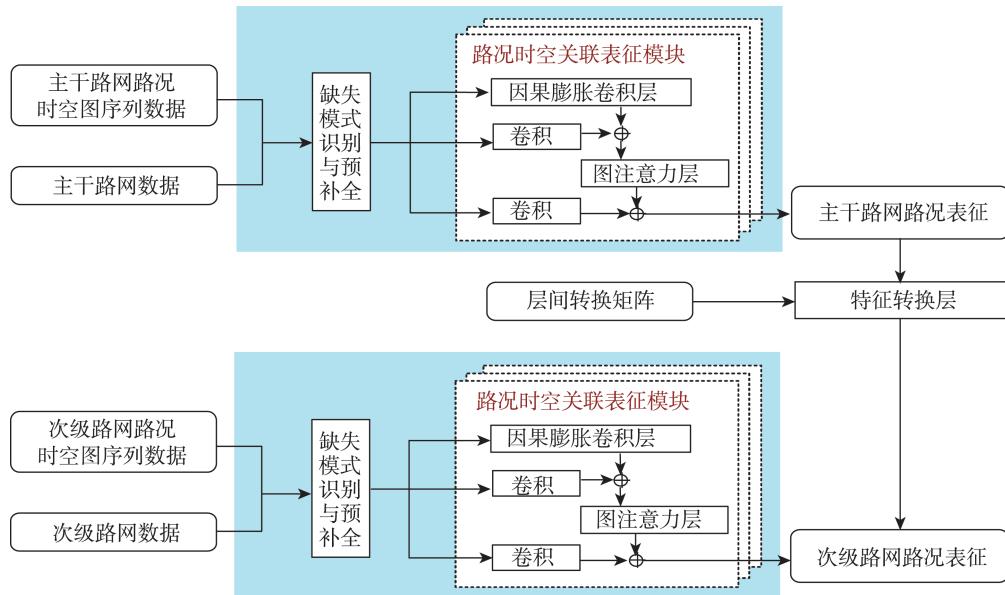


图4 基于分层路网的路况关联表征

Fig. 4 Correlation representation of traffic state based on the layered road network

2.5 多层时空图神经网络模型中的层内路况关联表征

T-MGNN 模型在不同层的路网上同时调用多层次路况时空关联表征模块。每个路况时空关联表征模块主要包括因果膨胀卷积层和图注意力层,公式如式(2)所示。

$$\begin{cases} \tilde{X} = CDCL(X_{t-p+1}^t) \\ \tilde{X}' = CDCL(\tilde{X}) + F^* X_{t-p+1}^t \\ O = GAL(\tilde{X}') + F^* X_{t-p+1}^t \end{cases} \quad (2)$$

式中: $X_{t-p+1}^t \in R^{n \times p}$ 表示预补全后的观测值; $\tilde{X}, \tilde{X}' \in R^{n \times p \times d}$ 表示计算过程的中间表征; n 表示该层级路段数量; p 表示输入的历史时段数; d 是特征维度,沿着空间维度 \tilde{X} 可以展开为 $\{\tilde{x}_i\}_{i=1}^n$, 用于建模特定层级路网的空间关系,沿着时间维度 \tilde{X} 可以展开为 $\{\tilde{x}^t\}_{t=t-p+1}^t$, 用于建模特定层级路网的时间关系; $CDCL$ (Causal Dilated Convolution Layer)表示因果膨胀卷积层; GAL (Graph Attentional Layer)表示图注意力层; F^* 表示 1×1 卷积运算,保留原始时空数据信息,路网数量较多时使模型更易收敛; O 表示 STCE 的输出结果。

因果卷积比传统卷积能更好地描述时间序列数据的依赖关系,更好地解决卷积层数增加带来的训练复杂和梯度消失等问题^[27]。因果膨胀卷积在因果卷积的基础上设定膨胀因子,不增加参数数量

同时扩大卷积的感受野,从而提升对时序特征的表征能力^[28]。本方法采用因果膨胀卷积表征路况时间维度上的关联:

$$F^* d X_{t-p+1}^t = \left(\sum_{i=1}^K w_i x^{(t-p-1)-(i-1)d} \dots \sum_{i=1}^K w_i x^{t-(i-1)d} \right) \quad (3)$$

式中: $F^* d X_{t-p+1}^t$ 表示对预补全后的观测值 $X_{t-p+1}^t \in R^{n \times p}$ 进行膨胀因子为 d 的因果膨胀卷积运算; $x^{t-(i-1)d} \in R^{n \times 1}$ 代表时刻 t 所有路段的预补全观测值; w_i 表示因果膨胀卷积中的可学习卷积核参数; K 表示卷积核大小。T-MGNN 可以调节因果膨胀卷积的卷积核大小 K 和膨胀因子 d 适应不同的时空序列输入长度,一般设置 $d=2^i$, i 代表第 i 层卷积。

受到图注意力网络^[29]的启发,在空间维度上,实现了图注意力层 GAL 为每个邻居路段分配不同权重,体现出路况在空间上的异质性,有效提取空间关联。

GAL 首先使用 Q 对拼接的中间状态升维增强特征表达,再计算节点 v_i 与领域中各节点的相关系数,归一化之后得到注意力系数(权重) γ_{ji} 。不同权重体现出路况信息在空间上的关联程度,权重越大的相邻路段表明其交通路况高度相关。最终使用注意力系数进行加权求和,聚合路况关联信息,得到每条路段 v_i 的路况空间特征 O_i :

$$\left\{ \begin{array}{l} O_i = \sum_{j \in A_i} \theta_{ji} \tilde{X}_i V \\ \theta_{ji} = \frac{\exp(\text{LeakyRelu}([\tilde{X}_i \| \tilde{X}_j'] Q))}{\sum_{k \in A_i} \exp(\text{LeakyRelu}([\tilde{X}_i \| \tilde{X}_k'] Q))} \end{array} \right. \quad (4)$$

式中: θ_{ji} 表示路段 v_j 对 v_i 的影响权重; $\tilde{X}_i' \in R^{n \times p \times e}$ 表示节点 i 的中间状态表征; \tilde{X}_j' 表示 i 的一个邻居节点 j 的中间状态表征; \tilde{X}_k' 表示用于执行求和操作以实现归一化的邻居节点集合中的元素; Q, V 表示 GAL 中可学习的参数; LeakyRelu 表示激活函数; \exp 表示指数函数; $[\cdot \| \cdot]$ 表示矩阵连接函数。

2.6 多层时空图神经网络模型中的层间消息传递

层间特征转换层(Trans)通过层间转换矩阵和层间有向注意力消息传递机制, 将主干路网提取的可靠路况时空信息逐级传递至轨迹稀疏的次级路网, 并与时空关联模块提取的次级路网路况时空特征融合。同时进行维度对齐, 输出表征各级路网路况关联的表征矢量, 输入到预测层进行路况预测(图5)。

以主干路网到次级路网的层间消息传递为例, 主干路网关联表征输出的特征 O_i^{cl} 与转换矩阵相乘进行升维操作, 使之与次级路网的特征 O_i^{c2} 维度对齐, 计算主干路网节点 v_i 传播至次级路网节点 v_j 的相关系数, 归一化后得到注意力系数(权重) η_{ij} , 使用注意力系数将主干路网特征传播至次级路网, 得到升维后的特征 $O_i'^{cl}$ 。具体计算公式如下:

$$\left\{ \begin{array}{l} O_i'^{cl} = \sum_{j \in A_i} \eta_{ji} O_i^{cl} SW \\ \eta_{ij} = \frac{\exp(\text{Relu}([O_i^{cl} S \| O_i^{c2}] W))}{\sum_{k \in S_i} \exp(\text{Relu}([O_i^{cl} S \| O_i^{c2}] W))} \end{array} \right. \quad (5)$$

式中: W 为可学习的参数; S 表示主干路网到次级路

网的转换矩阵。

路网分层关联表征模型整体前向传播公式如下:

$$\left\{ \begin{array}{l} O_{lt-p+1}^{c1:t} = \begin{cases} STCE(X_{t-p+1}^{c1:t}; W') & l=1 \\ STCE(O_{l-1:t-p+1}^{c1:t}; W') & 1 < l \leq N \end{cases} \\ O_{lt-p+1}^{c2:t} = \begin{cases} STCE(X_{t-p+1}^{c2:t}; W') & l=1 \\ Trans(O_{lt-p+1}^{c1:t}, \\ STCE(O_{l-1:t-p+1}^{c2:t}), W', S) & 1 < l \leq N \end{cases} \\ \hat{X}_{t+1}^{t+q} = F^* [O_{N,t-p+1}^{c1:t} \| O_{N,t-p+1}^{c2:t}] \end{array} \right. \quad (6)$$

式中: $X_{t-p+1}^{c1:t}$ 和 $X_{t-p+1}^{c2:t}$ 分别表示主干和次级路网预补全后的路况时序数据; $\hat{X}_{t+1}^{t+q} = \{\hat{x}\}_{t+1}^{t+q} \in R^{n \times q}$ 表示路网分层关联表征的输出, 即模型最终的预测结果, 其中 q 表示预测步长; $O_{lt-p+1}^{c1:t} \in R^{n \times p \times e}$ 表示主干路网(c1)建立的路况关联表征中第 l 个 STCE 模块的输出结果, $O_{lt-p+1}^{c2:t} \in R^{n \times p \times e}$ 表示次级路网(c2)建立的路况关联表征中第 l 个 STCE 组件的输出结果, 其中 e 表示路况时空数据在 STCE 模块中的特征维度; N 表示 STCE 模块的个数; F^* 表示一个卷积运算, 用于维度对齐; W' 表示可学习的参数。

2.7 多层时空图神经网络模型优化

路网分层关联表征模型建立后, 可以输入 p 步历史速度观测值 $X_{t-p}^t = \{x^r\}_{r=t-p}^t$, 通过预测计算层, 输出为未来 q 步路网的速度预测值 $\hat{X}_{t+1}^{t+q} = \{\hat{x}\}_{t+1}^{t+q}$ 。为了实现短期路况的预测, 在表征模型后增加速度-状态转换层, 对于每一个路段设置时间窗口为 5 的一维卷积核, 卷积核在时间维度滑动, 将路段速度映射成路段状态, 得到未来 q 步路况预测值 $\hat{X}_{con,t+1}^{t+q} = \{\hat{x}_{con}\}_{t+1}^{t+q} \in R^{n \times q}$ 。

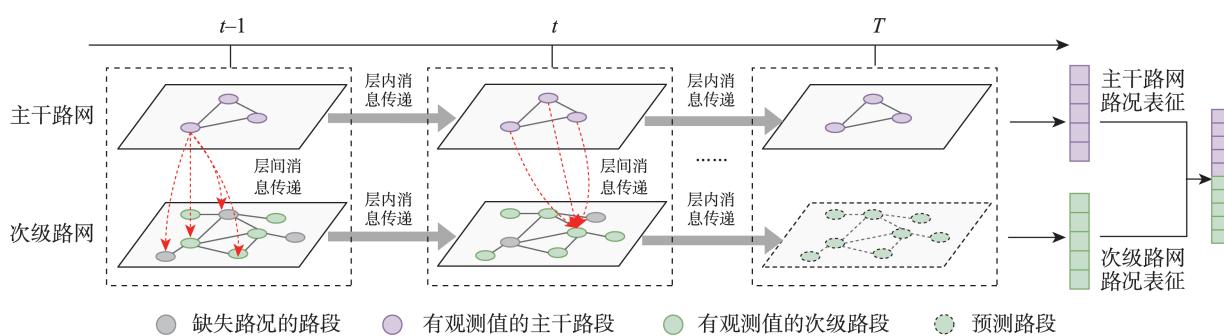


图 5 基于分层路网的层间和层内消息传递

Fig. 5 Inter-layer and intra-layer message passing based on layered road network

本文提出的损失函数由2部分组成:速度预测任务的损失和路况预测任务的损失。该损失用于同时保证速度和路况的预测精度,可以挖掘速度与路况之间的隐含关联,避免直接将速度转换为路况带来的精度损失。损失函数公式如式(7)所示。

$$L(\theta) = \min_{\theta} \left(\sum_{i=1}^q |\hat{x}_{con}^{t+i} - x_{con}^{t+i}| + \sum_{i=1}^q (\hat{x}^{t+i} - x^{t+i})^2 \right) \quad (7)$$

式中: x_{con}^{t+i} 为 $t+i$ 时刻所有路段的路况真值; x^{t+i} 为 $t+i$ 时刻所有路段的速度真值; $\sum_{i=1}^q (\hat{x}^{t+i} - x^{t+i})^2$ 表示预测速度的L2损失; $\sum_{i=1}^q |\hat{x}_{con}^{t+i} - x_{con}^{t+i}|$ 表示预测路况的L1损失; θ 表示模型中可学习的参数。

3 实验与结果分析

3.1 实验准备

实验所需要的数据包括路网和轨迹数据2类。

实验选择轨迹数据较为丰富的武汉市三环线以内及沌口、吴家山、光谷东3个片区的路网(图6)。路网由Open Street Map开放平台获取,并利用QGIS软件进行简化。简化后路网含路段边5 659条,交叉点4 419个。选取2018年7月2日—7月8日一周内连续运行的出租车轨迹数据进行实验,单日轨迹包含车辆约1.243万辆,采样平均间隔为10 s,采样频率不低于0.2 Hz。使用结合隐马尔可夫模型和预计算的快速地图匹配算法对轨迹数据进行地图匹配^[30],确定各轨迹点所属路段。

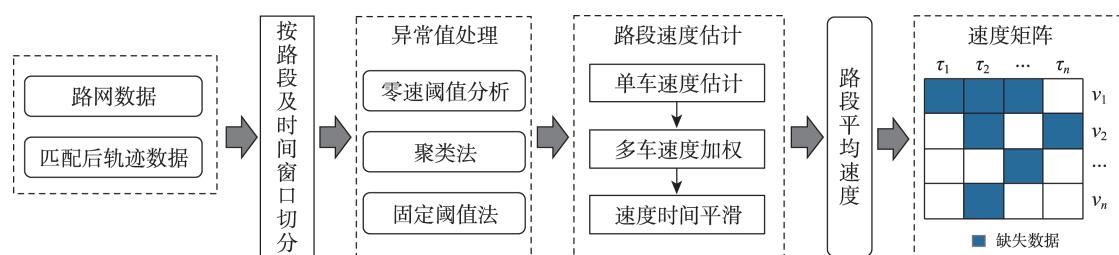
采取了原始轨迹数据后,需要基于轨迹速度进行预处理。基于分级后的路网,面向每个路段,去除速度异常的轨迹点,占比4.89%。设计实现速度提取算法计算每条路段每个时间窗口(如2~5 min)的平均速度,如图7所示。

轨迹数据中存在较多异常数据,如速度小于0或者大于120、在一定时间内距离不发生变化、长期停车等情况,在进行速度估计之前需要对异常数据



图6 研究区域的路网

Fig. 6 The road network of the study region



注: v_1, \dots, v_n 表示路段; τ_1, \dots, τ_n 表示划分的时间窗口。

图7 速度提取流程

Fig. 7 Speed extraction process

进行清洗。零速阈值分析法指判断并剔除处在停车状态的零速轨迹点;聚类法指对同一道路上的轨迹点进行聚类,剔除同一聚类中轨迹点速度与其他轨迹点相差较大的数据;固定阈值法指剔除速度在阈值范围外的轨迹点。

路段速度估计流程包括单车速度估计、多车速度加权、速度时间平滑3部分。首先根据出租车在当前路段以及前后路段交汇处的轨迹点估计单辆车的速度,其次根据GPS轨迹数据质量对多辆车的速度进行加权计算得到路段速度^[31],最后根据路段轨迹点数量的对速度进行时间平滑。

速度提取完成后得到速度矩阵,其中横轴表示时间维度,纵轴表示空间维度,即每一行为一条路段所有时间窗口的观测值,每一列为一个时间窗口所有路段的观测值。速度矩阵每个时间窗口大小为5 min,覆盖了5 659条路段的一周观测数据。

选择工作日7—9时以及17—19时高峰期数据,统计各路段5 min窗口的平均流量,如图8所示。流量大于6的路段设置为主干路网,共1 598条,占全部路网的28%,其余路段设置为次级路网,共4 061条,占全部路网的72%。

武汉数据集中所有路段均存在不同程度的缺失,在保证路段连通性条件下去除缺失率大于60%的路段837条,得到包含4 822条路段的路网时空数据进行训练,其中主干路段1 446条,次级路段3 376条。时空序列数据由速度观测值和路况观测值组成,均包含9 721 152个样本,其中速度观测值使用Z-score归一化后输入。数据集取每天的7—22时,按照8:2比例在时间维度划分训练集和测试集。

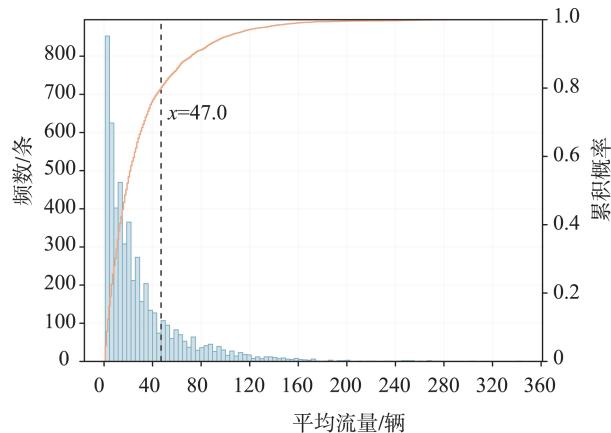


图8 武汉市2023年8月某日路段5 min流量频数直方图

Fig. 8 Traffic volume Histogram for roads in Wuhan on a certain day in August 2023 (5-minute interval)

3.2 实验设置

采用4种预测指标来衡量不同模型性能,对于速度预测任务,使用平均绝对误差(Mean Absolute Error, MAE)、均方根误差(Root Mean Squared Error, RMSE)和平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)。计算公式如式(8)—式(10)所示。

$$MAE = \frac{1}{\Omega} \sum_{i=1}^{\Omega} |x_i - \hat{x}_i| \quad (8)$$

$$RMSE = \sqrt{\frac{1}{\Omega} \sum_{i=1}^{\Omega} (x_i - \hat{x}_i)^2} \quad (9)$$

$$MAPE = \frac{1}{\Omega} \sum_{i=1}^{\Omega} \left| \frac{x_i - \hat{x}_i}{x_i} \right| \times 100\% \quad (10)$$

式中: Ω 为样本数量,某一时刻某一路段的观测值为一个样本; x_i 为观测值的真值; \hat{x}_i 为T-MGNN模型的输出预测值。

对于交通拥堵状态预测任务,使用状态一致率指标。计算公式如式(11)所示。

$$Consistency = \frac{\sum_{i=1}^{\Omega} len(x_i) \times (x_i \cap \hat{x}_i)}{\sum_{i=1}^{\Omega} len(x_i) \times (x_i \cup \hat{x}_i)} \quad (11)$$

式中: x_i 为拥堵状态真值; \hat{x}_i 为预测值;0代表畅通;1代表拥堵; $len(x_i)$ 表示当前样本的路段长度,分子表示预测和真值均为拥堵所有样本的路段长度之和,分母表示预测或真值为拥堵所有样本的路段长度之和。

本文实验基于Python 3.9和PyTorch软件包。模型实验均用9个历史时间步的时空数据预测未来4个时间步的时空数据,T-MGNN模型因果卷积层由2层卷积组成,每层卷积核大小为2,膨胀因子设置为2,4,主干路网和次级路网均具有2个路况时空关联表征模块,使用Adam优化器训练模型,初始学习率为0.001,批大小为16,训练周期数为1 000。

3.3 实验结果分析

本节选取近年来浅层机器学习模型、基于图的深度学习模型和基于注意力的深度学习模型3大类中表现较为优异,代码开源可复现的8个基线模型与本文提出的T-MGNN模型进行比预测精度指标对比。由于大规模路网路段数量较多,存在部分轨迹极少路段或异常值,仅选取轨迹数量大于5的路段参与精度指标计算。全路网指筛选后路段中所有主干路段和次级路段,全路网精度指标结果如表1所

表1 T-MGNN与基线模型全路网预测的精度对比

Tab. 1 Comparison of T-MGNN and baseline models on the prediction performance for the entire road network

模型	速度						拥堵状态	
	单步预测 5 min			四步预测 20 min			单步预测 5 min	四步预测 20 min
	MAE/(km/h)	RMSE/(km/h)	MAPE/%	MAE/(km/h)	RMSE/(km/h)	MAPE/%	状态一致率/%	
ST-KNN ^[32]	2.80	4.21	9.51	4.73	6.33	14.10	78.28	67.74
BTMF ^[33]	3.26	4.97	10.24	3.97	5.28	11.12	81.62	78.98
SGMN ^[26]	2.18	3.13	6.04	3.86	5.62	11.45	85.52	79.44
T-GCN ^[13]	3.48	5.19	11.90	3.71	5.42	12.26	73.87	72.59
DGCNN ^[18]	2.06	2.95	7.89	4.21	5.90	15.92	85.68	70.64
PDFormer ^[19]	2.09	3.48	6.54	3.64	5.17	10.88	85.78	82.26
Causal-GCN ^[20]	2.26	3.33	7.19	3.78	5.56	11.67	83.25	77.10
D-TGNM ^[24]	2.25	3.39	7.30	3.69	5.31	10.99	83.44	79.23
T-MGNN	2.01	2.91	5.45	3.26	4.69	9.82	86.18	84.35

示。深度学习方法优于浅层机器模型方法，在路况预测中具有更强的学习有效特征的能力。基于图的深度学习模型中，使用动态图的模型预测精度显著高于 SGMN、T-GCN，可能是因为动态图也能学习部分大规模路网中的层次结构，相对于静态图能更好地表征路网之间地拓扑关系。T-MGNN 采用了动态路网分层方法和路网分层关联表征模型，更好地捕捉到大规模路网层次结构，而不是仅依赖于邻接矩阵获取路况动态变化信息，因此预测性能上表现最佳。预测时间步长由单步增加至四步后，所有模型的预测精度都出现了不同程度的下降，其中 DGCNN 衰减幅度最大，而 T-MGNN 仍然能保持较好预测性能，这可能是因为动态循环卷积在大规模路网上更不易收敛，同时也说明短期预测中 STCE 模块在时间维度上提取特征的有效性。状态一致率的变化趋势和 MAPE 高度相关，且比 MAPE 更为敏

感，因为预测速度值直接影响到拥堵状态，计算指标时加入路段长度放大了误差造成的影响，因此也更能反映模型对稀疏轨迹路段的预测水平，T-MGNN 在表征-预测一体化基础上协同速度和拥堵状态关系，因此 MAPE 和状态一致率均表现最优秀。总体来说，T-MGNN 在全路网的单步预测和四步预测均取得了最好的预测性能。

本文对筛选后路网中次级路网进一步分析，结果如表2所示。由于大规模城市路网轨迹分布存在不均衡性，次级路网比主干路网轨迹稀疏，路况变化频繁更为复杂，各路段速度更不可靠，所有模型在次级路网上计算的指标均下降。基于动态图的模型和基于注意力的模型下降幅度小于浅层机器学习模型和基于静态图的模型，得益于分层结构中的层间消息传递机制，T-MGNN 将主干路网中提取的路况关联表征转换至次级路网加以利用，有效提

表2 T-MGNN与基线模型次级路网预测的精度对比

Tab. 2 Comparison of T-MGNN and baseline model on the prediction performance for the secondary road network

模型	速度						拥堵状态	
	单步预测 5 min			四步预测 20 min			单步预测 5 min	四步预测 20 min
	MAE/(km/h)	RMSE/(km/h)	MAPE/%	MAE/(km/h)	RMSE/(km/h)	MAPE/%	状态一致率/%	
ST-KNN ^[32]	4.76	6.32	14.64	4.81	6.46	14.95	57.70	56.70
BTMF ^[33]	3.74	5.23	10.98	4.73	5.94	12.22	73.47	67.70
SGMN ^[26]	2.78	3.55	7.01	4.01	5.65	11.62	77.11	69.21
T-GCN ^[13]	3.57	5.58	12.03	4.14	5.78	12.47	67.64	66.61
DGCNN ^[18]	2.11	3.39	8.10	4.57	6.13	16.54	79.87	63.30
PDFormer ^[19]	2.16	3.61	6.82	3.70	5.39	10.48	80.23	70.57
Causal-GCN ^[20]	2.48	3.47	8.02	3.88	5.67	11.71	75.30	67.23
D-TGNM ^[24]	2.50	3.55	7.98	3.86	5.60	11.25	78.75	69.45
T-MGNN	2.02	3.14	5.84	3.40	5.02	10.11	81.44	72.36

升了轨迹稀疏路段的路况表征质量,因此依然能够达到最佳预测水平。单步预测和四步预测的变化情况与全路网类似,但是次级路网中二者的差异更大,MAPE 和状态一致率的衰减更为剧烈,T-MGNN 中的分层结构和速度-状态转换层能够充分利用主次路网的结构信息和路况信息,相较于基线模型中的单层图神经网络,能够利用高级邻域中提取到远程的时空关联信息进行消息传递,进而挖掘次级路网时空特征。综上所述,T-MGNN 在全路网和轨迹数据缺失较严重的次级路网上均具有明显优势。

本文数据集包含 4 822 条路段,使用该大规模路网时空数据训练深度学习模型时,相较于小型数据集,对模型显存需求更大,且训练时间更长,因此训练效率的评估也非常重要。选取深度学习模型中表现良好的 DGCRN、PDFomer 和 Causal-GCNM 模型,并增加 T-MGNN 的分层结构和不分层结构进行对比。设置相同环境配置和数据集批大小,分别计算 1 个周期训练耗时和测试时 1 次迭代的推理时间。为了确保数据准确,从第 20 个训练周期开始,随机取 10 个单周期训练耗时计算平均值作为训练时间,重复 5 次测试过程取迭代平均耗时作为推理时间。结果如表 3 所示。

DGCRN 训练和推理均最慢,可能的原因是使用了循环卷积模块,同时需要重新生成动态邻接矩阵。Causal-GCNM 在时间维度上使用空间卷积模块,因此效率相较于同样使用图神经网络的 DGCRN 大幅提高。基于注意力机制的 PDFomer 效率较高,但是在训练过程中显存占用非常大,这可能是路段数量庞大导致矩阵运算占用过多空间导致的。未分层的 T-MGNN 表现一般,但是采用分层结构后训练提速接近 7 倍,同样大幅领先其他深度学习模型。本文推理时间使用单次迭代计算,数量级为毫秒,因此差距较小,T-MGNN 凭借其较小的参数量获得了微弱优势。

表 3 各模型 4 步预测训练和推理时间

Tab. 3 Training and inference time of the four-step prediction of all compared models

模型	训练时间/epoch (s)	推理时间/iter (ms)
SGMN ^[26]	33.55	6.32
T-GCN ^[13]	35.76	4.78
DGCRN ^[18]	85.42	105.06
PDFomer ^[19]	34.03	4.63
Causal-GCNM ^[20]	38.38	5.27
T-MGNN(未分层)	37.35	5.11
T-MGNN(分层)	5.53	3.91

未分层 T-MGNN 和分层后 T-MGNN 训练时精度指标变化如图 9 所示。同样的训练周期数,分层后的 T-MGNN 能更快收敛,且每个周期耗时更短,这显示了分层结构对模型整体信息传递的有效性,在大规模路网路况预测方面具备很大潜力。

交通路况随时间变化波动,对 2018 年 7 月 2 日 9:14 和 20 时这 3 个不同时段的单步预测结果进行评估,选取表现较好的 PDFomer、Causal-GCNM 和 T-MGNN 模型进行对比。如图 10 所示,PDFomer 在 MAE、MAPE 指标上表现优于 Causal-GCNM。3 个模型在 14:00 的预测精度均降低,该为非高峰期,轨迹数量较少,提取的速度相较于高峰期 9:00 和 20:00 更不可靠,说明轨迹密度会直接影响到模型预测。T-MGNN 在 3 个时段所有指标均取得最好表现,再次证明具备分层结构的 T-MGNN 能提升稀疏轨迹路段的预测性能,在大规模路网上拥有更强的模拟复杂多变路况能力。

以 9:00 时段为例,对 PDFomer、Causal-GCNM 和 T-MGNN 进行全路网的路况分析,如图 11 所示,图中红色代表拥堵,蓝色代表畅通。对于城市主干路网的拥堵状态,3 个模型都具有良好的预测性能,但是在一些轨迹稀疏的城市边缘区域,如三环线外的沌口和光谷东,PDFomer 和 Causal-GCNM 与真值均出现较多偏差,而 T-MGNN 依然能正确进行路况预测。武汉二环线内的一些轨迹密集区域,如江汉路和徐东,存在很多长度较短的次级路网,因此路况非常复杂,PDFomer 整体预测分布均匀,但是出现了很多非拥堵路况预测成为拥堵路况的情况。Causal-GCNM 倾向于低估拥堵路况,且大部分长度较小路段的拥堵状况未能预测到,这些路段大部分

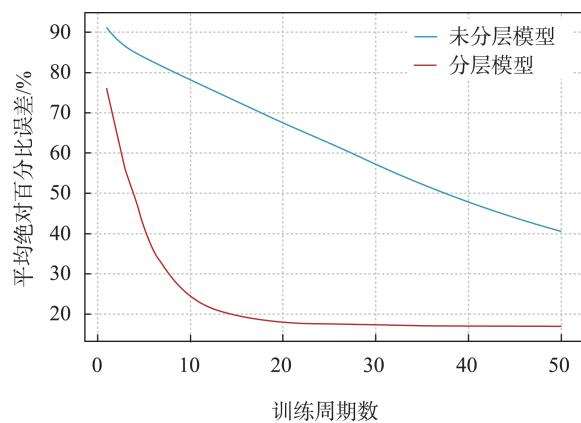


图 9 T-MGNN 分层与未分层训练过程对比

Fig. 9 Comparison of the training processes of T-MGNN on layered and original road networks

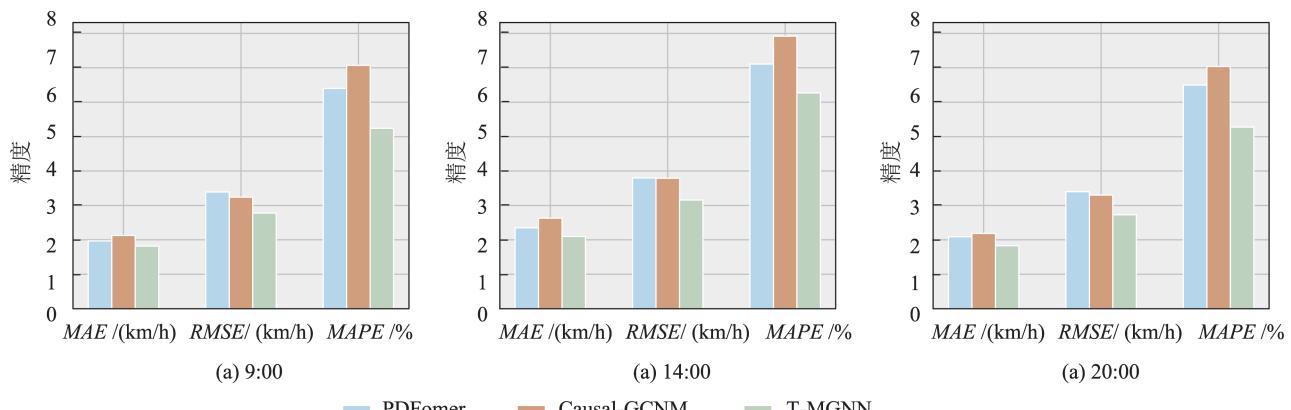


图 10 PDFormer、Causal-GCN 和 T-MGNN 不同时段的精度对比

Fig. 10 Comparison of prediction performance on different time periods of PDFormer, Causal GCNM, and T-MGNN

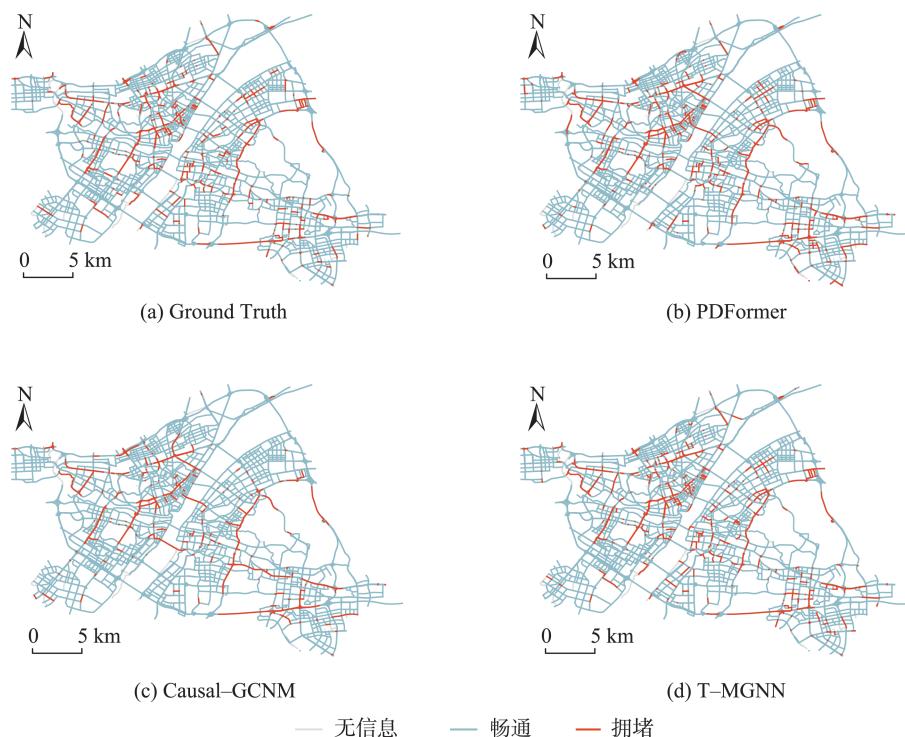


图 11 拥堵状况预测结果

Fig. 11 Congestion prediction results

属于次级路网,说明主干路网和次级路网之间的交通信息流通效率较低。T-MGNN 预测路况分布均匀,预测拥堵状况与真值接近,在两级路网均取得最佳表现,证明 T-MGNN 能将主干路网信息充分利用至次级路网,在更小的颗粒度上也能实现精确的拥堵路况预测。

为了更为直观地观察 3 个模型在图 11 预测结果中情况,将预测结果与真值对比后进行分析,结果如图 12 所示。错误路段包括将非拥堵预测成拥

堵和拥堵预测成非拥堵 2 种情况。PDFormer 和 Causal-GCN 存在较多错误预测情况,T-MGNN 错误预测路段仅包含细碎路段,这可能是由于这些路段轨迹数量过少,提取的速度值不可靠,且周围没有主干路网传递交通信息,因此无法正确预测拥堵。

在武汉数据集上对 T-MGNN 模型进行消融实验,用以验证 T-MGNN 分层结构和关键模块的有效性,结果见表 4。w/o STCE 代表使用普通因果卷积和图卷积代替 STCE 模块的模型。w/o hier 为

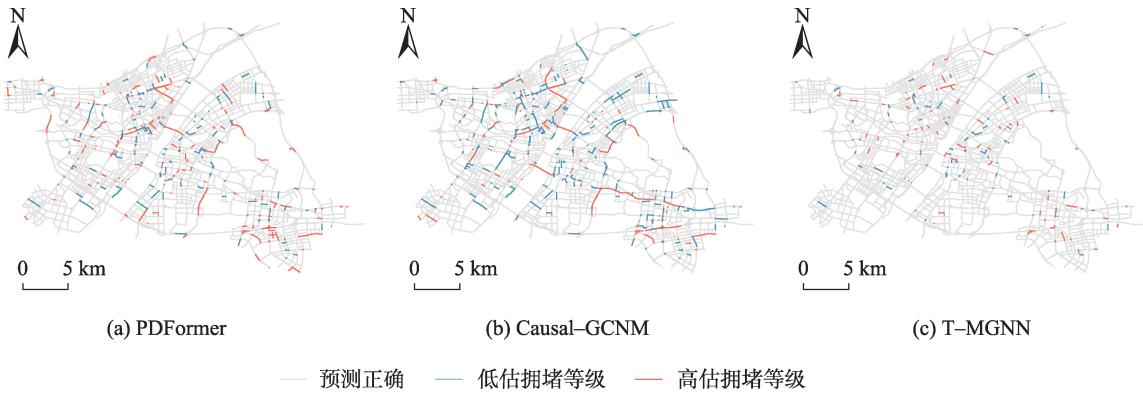


图 12 拥堵预测结果与真值对比

Fig. 12 Comparison between congestion prediction results and true values

表4 不同组件对T-MGNN影响

Tab. 4 The impact of different components on T-MGNN

模型	单步预测												四步预测											
	全路网						次级路网						全路网						次级路网					
	MAE /(km/h)	RMSE /(km/h)	MAPE /%	状态 一致率																				
w/o STCE	4.68	6.76	12.30	71.66	4.91	6.90	12.95	68.85	4.86	6.15	14.17	70.41	5.57	6.73	15.37	65.01								
w/o hier	3.07	4.22	8.67	80.26	3.90	5.37	9.78	70.75	4.25	5.96	11.83	76.34	4.70	6.21	12.90	65.17								
w/o cond	2.55	3.46	7.86	81.19	2.67	3.70	7.91	77.63	3.57	5.29	10.49	80.58	3.94	5.83	10.85	67.01								
w/ concat	2.49	3.40	7.83	84.98	2.56	3.73	7.94	79.46	3.54	5.33	10.52	82.68	3.78	5.95	11.06	68.13								
T-MGNN	2.01	2.91	5.45	86.18	2.02	3.14	5.84	81.44	3.26	4.69	9.82	84.35	3.40	5.02	10.11	72.36								

没有分层结构的 T-MGNN, STCE 模块串联进行训练和测试。w/o cond 在 T-MGNN 基础上去除了速度-状态转换层, 仅用拥堵阈值将路段预测速度映射至路段状态。w/ concat 表示仅拼接主干路网特征和次级路网特征, 不使用层间特征转换层。使用 STCE 模块可以提升所有精度指标, 表明 STCE 能捕获路网交通状况中的复杂的时空关联。使用分层结构后, 模型整体预测性能提升, 同时次级路网预测误差和主干路网预测误差差距减小, 这说明层间消息传递能充分发掘大规模路网中蕴含的层次结构, 表明路况预测中引入分层结构的必要性和有效性。增加了速度-状态转换层后, 预测性能进一步提升, 且状态一致率提升幅度较大, 原因是模型学习到速度和状态之间的相关性, 而不是简单地直接根据拥堵阈值进行转换。层间消息传递时使用的特征转换层相较于直接拼接可以获得更好的精度指标, 说明主干路网中道路对周边的次级路网中各路段影响存在差异, 特征转换层中的注意力机制能赋予不同权重, 从而更精准地在主次路网之间进行信息流通。

4 结论与展望

本文创新了一种依托路网分层的多层次时空图神经网络交通路况预测方法, 可以直接基于大规模出租车轨迹数据进行大型路网路况的短期预测。通过武汉市实际轨迹数据实验证明, 本方法比 8 种流行的基线方法在路况预测精度上有了显著改善, 特别是在轨迹数据缺失的路段上在速度和状态预测精度方面都优于基线方法。所提出的预测方法在训练和推理效率方面也有较大的优势, 特别是路网分层提高了模型训练效率。通过实验表明所提出的多层次时空图神经网络预测方法能有效地应对轨迹分布不均衡导致的路况预测难题。本研究的主要创新和贡献总结如下:

(1) 基于大规模原始轨迹数据的动态路网分层方法。根据轨迹体现的交通模式在不同时段上对路网进行分层, 分解为主干和若干级次要路网, 从而更好地分解轨迹数据的时空不均衡问题, 为建立多层次时空图神经网络建立基础数据结构;

(2) 基于多层次时空图神经网络的路网分层关联

表征模型。依托分层路网,提出基于因果膨胀卷积和图注意力的多层时空图神经网络,设计层内和层间消息传递机制,更好地描述路网之间复杂的路网时空关联;

(3)表征-预测一体化集成短期预测建模方法架构。在路况关联表征模型基础上,设计实现了表征-预测一体化的集成图神经网络预测模型,可同时对全路网所有路段的速度和状态进行在线预测,有效提升轨迹分布稀疏路段的路况预测质量。

本文方法对于提升城市交通路况的监测精度和效率,改进交通在线控制和组织,支持出行者信息服务等上层应用都有着一定的应用价值。本文方法提供了完善实用化的技术流程,可以直接用于交通管理的实际应用。在潜在的缺点方面,方法在应对轨迹极度稀缺时可能无法构建较为完整的主干路网。此时,不同层级路网的消息传递的效果也会有所下降。因此未来将探讨最少需要多少轨迹数据才能发挥模型的性能。此外受到数据可得性限制,本文方法仅在武汉市数据集上开展了测试,未来研究将在更多的城市应用和改进本方法。同时进一步改善路网自适应分层的方法,更好地应对轨迹的时空分布不平衡。在路况关联表征方面,可以尝试应用图 Transformer 框架并改进层间消息传递机制,进一步提升次级路网的路况预测性能。

致谢:本文的数值计算得到了武汉大学超级计算中心的计算支持和帮助。

参考文献(References):

- [1] Zhang L M, Zeng G W, Li D Q, et al. Scale-free resilience of real traffic jams[J]. Proceedings of the National Academy of Sciences of the United States of America, 2019, 116(18):8673-8678. DOI:10.1073/pnas.1814982116
- [2] Rempe F, Franeck P, Bogenberger K. On the estimation of traffic speeds with deep convolutional neural networks given probe data[J]. Transportation Research Part C: Emerging Technologies, 2022, 134:103448. DOI:10.1016/j.trc.2021.103448
- [3] 朱秋圳,邬群勇,姚铖鑫,等.基于DBI和稀疏轨迹数据的交通状态精细划分与识别[J].地球信息科学学报,2022, 24(3):458-468. [Zhu Q Z, Wu Q Y, Yao C X, et al. Fine classification and identification of traffic states based on DBI and sparse trajectory data[J]. Journal of Geo-information Science, 2022, 24(3):458-468.] DOI: 10.12082/dqxxk x.2022.210408
- [4] Tsubota T, Bhaskar A, Nantes A, et al. Comparative analysis of traffic state estimation[J]. Transportation Record: Journal of the Transportation Research Board, 2015, 2491(1):43-52. DOI:10.3141/2491-05
- [5] Shaygan M, Meese C, Li W X, et al. Traffic prediction using artificial intelligence: review of recent advances and emerging opportunities[J]. Transportation Research Part C: Emerging Technologies, 2022, 145:103921. DOI:10.1016/j.trc.2022.103921
- [6] 姚俊峰,何瑞,史童童,等.基于机器学习的交通流预测方法综述[J].交通运输工程学报,2023,23(3):44-67. [Yao J F, He R, Shi T T, et al. Review on machine learning-based traffic flow prediction methods[J]. Journal of Traffic and Transportation Engineering, 2023, 23(3): 44-67.] DOI:10.19818/j.cnki.1671-1637.2023.03.003
- [7] Rahmani S, Baghbani A, Bouguila N, et al. Graph neural networks for intelligent transportation systems: A survey [J]. IEEE Transactions on Intelligent Transportation Systems, 2023,24(8):8846-8855. DOI: 10.1109/TITS.2023.3257759.
- [8] Li Y G, Yu R, Shahabi C, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting[EB/OL]. 2017:1707.01926.<http://arxiv.org/abs/1707.01926v3>
- [9] Ren Y B, Chen H F, Han Y, et al. A hybrid integrated deep learning model for the prediction of citywide spatio-temporal flow volumes[J]. International Journal of Geographical Information Science, 2020, 34(4): 802-823. DOI: 10.1080/13658816.2019.1652303
- [10] Jiang W W, Luo J Y. Graph neural network for traffic forecasting: A survey[J]. Expert Systems with Applications, 2022,207:117921. DOI:10.1016/j.eswa.2022.117921
- [11] Yu B, Yin H T, Zhu Z X. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting[C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2018:3634-3640. DOI:10.24963/ijcai.2018/505
- [12] Wu Z H, Pan S R, Long G D, et al. Graph wavenet for deep spatial-temporal graph modeling[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence. ACM, 2019:1907-1913. DOI:10.5555/3367243.3367303
- [13] Zhao L, Song Y J, Zhang C, et al. T-GCN: A temporal graph convolutional network for traffic prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2020,21(9):3848-3858. DOI: 10.1109/TITS.2019.2935152
- [14] Zhang Y, Cheng T, Ren Y B, et al. A novel residual graph convolution deep learning model for short-term network-

- based traffic forecasting[J]. International Journal of Geographical Information Science, 2020,34(5):969-995. DOI: 10.1080/13658816.2019.1697879
- [15] Li Q M, Han Z C, Wu X M. Deeper insights into graph convolutional networks for semi-supervised learning[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2018,32(1):3538-3545. DOI:10.1609/aaai.v32i 1.11604
- [16] Ta X X, Liu Z H, Hu X, et al. Adaptive spatio-temporal graph neural network for traffic forecasting[J]. Knowledge-Based Systems, 2022,242:108199. DOI:10.1016/j.knosys.2022.108199
- [17] Guo K, Hu Y L, Sun Y F, et al. Hierarchical graph convolution network for traffic forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2021,35(1):151-159. DOI:10.1609/aaai.v35i1.16088
- [18] Li F X, Feng J, Yan H, et al. Dynamic graph convolutional recurrent network for traffic prediction: Benchmark and solution[J]. ACM Transactions on Knowledge Discovery from Data, 2023,17(1):9. DOI:10.1145/3532611
- [19] Jiang J W, Han C K, Zhao W X, et al. PDFormer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2023,37(4):4365-4373. DOI:10.1609/aaai.v37i4.25556
- [20] 王培晓,张彤,聂士超,等.顾及缺失值的因果图时空预测网络[J].测绘学报,2023,52(5): 818-830. [Wang P X, Zhang T, NIE S C, et al. A causal graph convolutional network considering missing values for spatio-temporal prediction[J]. Acta Geodaetica et Cartographica Sinica, 2023, 52(5):818-830.] DOI:10.11947/j.AGCS.2023.20220021
- [21] Yang J M, Peng Z R, Lin L. Real-time spatiotemporal prediction and imputation of traffic status based on LSTM and Graph Laplacian regularized matrix factorization[J]. Transportation Research Part C: Emerging Technologies, 2021,129:103228. DOI:10.1016/j.trc.2021.103228
- [22] Wang P X, Zhang T, Zheng Y M, et al. A multi-view bidirectional spatiotemporal graph network for urban traffic flow imputation[J]. International Journal of Geographical Information Science, 2022,36(6):1231-1257. DOI:10.1080/13658816.2022.2032081
- [23] Tian Y, Zhang K L, Li J Y, et al. LSTM-based traffic flow prediction with missing data[J]. Neurocomputing, 2018, 318:297-305. DOI:10.1016/j.neucom.2018.08.067
- [24] Wang P X, Zhang Y, Hu T, et al. Urban traffic flow prediction: A dynamic temporal graph network considering missing values[J]. International Journal of Geographical Information Science, 2023,37(4):885-912. DOI:10.1080/13658816.2022.2146120
- [25] Wang A, Ye Y C, Song X Z, et al. Traffic prediction with missing data: A multi-task learning approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(4):4189-4202. DOI:10.1109/TITS.2022.3233890
- [26] Cui Z Y, Lin L F, Pu Z Y, et al. Graph Markov network for traffic forecasting with missing data[J]. Transportation Research Part C: Emerging Technologies, 2020, 117: 102671. DOI:10.1016/j.trc.2020.102671
- [27] Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016,39(4): 640-651. DOI:10.1109/TPAMI.2016.2572683
- [28] Van Den Oord A, Dieleman S, Zen H G, et al. WaveNet: A generative model for raw audio[EB/OL]. 2016: 1609.03499.<http://arxiv.org/abs/1609.03499v2>
- [29] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[EB/OL]. 2017:1710.10903,2017.<https://arxiv.org/abs/1710.10903>
- [30] Yang C, Gidófalvi G. Fast map matching, an algorithm integrating hidden Markov model with precomputation[J]. International Journal of Geographical Information Science, 2018,32(3):547-570. DOI:10.1080/13658816.2017.1400548
- [31] 张继仙.基于GPS数据质量评估的路段速度估计方法研究[D].广州:中山大学,2012.[Zhang J X.Research of link speed estimation based on traffic GPS data quality analysis[D]. Guangzhou: Sun Yat-sen University, 2012.]
- [32] Yu B, Song X L, Guan F, et al. K-nearest neighbor model for multiple-time-step prediction of short-term traffic condition[J]. Journal of Transportation Engineering, 2016,142 (6):04016018. DOI:10.1061/(asce)te.1943-5436.0000816
- [33] Chen X Y, Sun L J. Bayesian temporal factorization for multidimensional time series prediction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(9):4659-4673. DOI:10.1109/TPAMI.2021.3066551